

# Overview of Procedural Approach for Migrating Geologic Map Data and Related Processes to a Geodatabase

By James R. Chappell<sup>1</sup>, Stephanie A. O'Meara<sup>1</sup>, Heather S. Stanton<sup>1</sup>,  
Greg Mack<sup>2</sup>, Anne R. Poole<sup>3</sup> and Georgia Hybels<sup>4</sup>

<sup>1</sup>Colorado State University/National Park Service Cooperator

Colorado State University

Fort Collins, CO 80521

Telephone: (970) 491-5147

e-mail: Jim\_Chappell@partner.nps.gov; Stephanie\_O'Meara@partner.nps.gov;  
Heather\_Stanton@partner.nps.gov

<sup>2</sup>National Park Service – Pacific West Region, Seattle Office

e-mail: Greg\_Mack@nps.gov

<sup>3</sup>United States Forest Service – Chippewa National Forest

e-mail: apoole@fs.fed.us

<sup>4</sup>University of Denver/National Park Service Cooperator

e-mail: Georgia\_Hybels@partner.nps.gov

## INTRODUCTION

Geologic maps are an integral component of the physical science inventories stipulated by the National Park Service (NPS) in its Natural Resources Inventory and Monitoring (I&M) Guideline (<http://science.nature.nps.gov/im/index.cfm>). The NPS Geologic Resources Division (GRD) is currently developing a Geologic Resources Evaluation (GRE) that includes a geologic bibliography, the creation of summary reports of each park's geology, evaluation of existing geologic maps, and the development of a geology-GIS data model for implementation in the production of digital geologic-GIS data for each park (such as Rocky Mountain National Park or Great Sand Dunes National Park and Preserve). The current data model implemented by the GRE for digital geologic-GIS data is the NPS GRE Geology-GIS Coverage/Shapefile Data Model (O'Meara et. al., 2005).

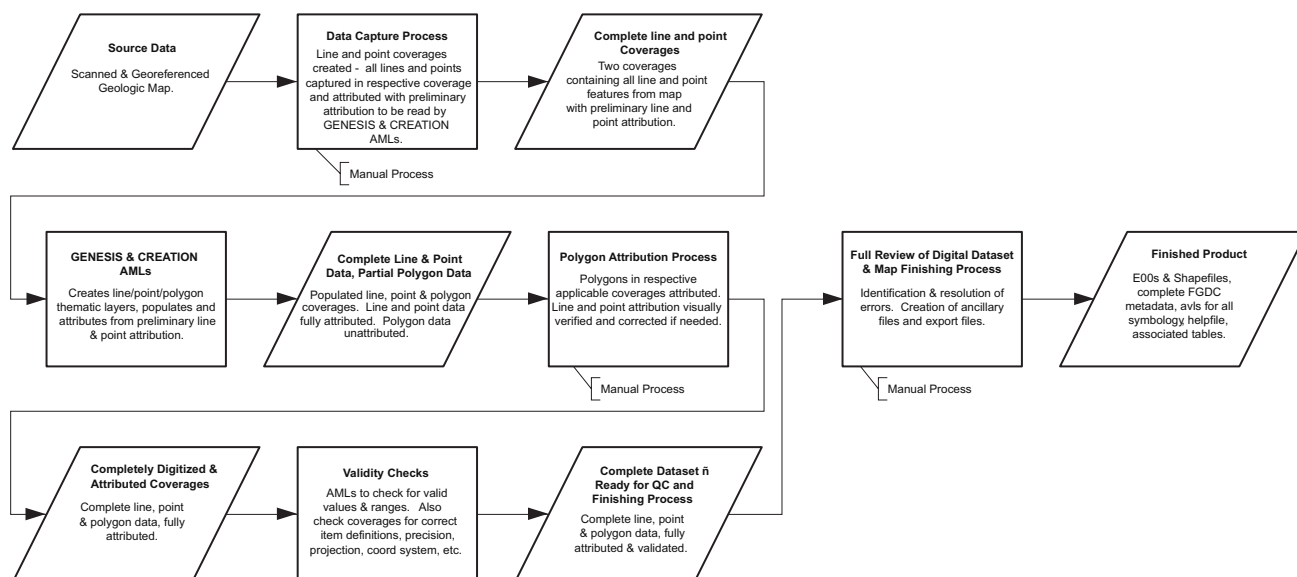
Recently, Environmental Systems Research Institute (ESRI) released the personal geodatabase, a relational database management system (RDBMS) designed specifically for storing, updating and viewing spatial data. Compared to the coverage- and Shapefile-based GIS previously offered by ESRI, the personal geodatabase offers added functionality of attribute validation rules, relationship classes, and topological rules that maintain data integrity within and between data layers. The GRE has determined that the added functionality of the geodatabase will

increase data quality and help stream-line the data production process. Migration of GIS data involves significant changes to the current GRE data model and the revision of existing data capture/conversion procedures. Additionally, the migration of GRE legacy data must also be addressed.

## CURRENT DATA FORMATS AND PRODUCTION PROCESSES

Presently, all GRE digital geologic-GIS data are stored in both coverage and shapefile format. Completed digital geologic-GIS data for a specific park are comprised of a set of both shapefiles and coverages, with each set being a collection of data layers such as geologic contacts or faults, as defined by the NPS Geology-GIS Coverage/Shapefile Data Model (O'Meara et. al., 2005). The data layers included in each set can vary depending on the source maps from which the data was derived.

At present, the data capture process involves either hand digitization of hard-copy geologic maps or, less commonly, conversion of existing digital data. Digitization and conversion are primarily conducted using ESRI ArcInfo Workstation and ArcView 3.x software. The multi-step, modularized process (Figure 1) is segmented by Arc Macro Language (AML) scripts that aid in capture/conversion steps, provide some quality control and, most importantly, preserve topological relationships between geologic features on the map.



**Figure 1.** Schematic showing NPS GRE modular process and workflow for digitizing GIS data into the coverage/shapefile data model. Data is digitized and attributed in two datasets, one for point features, and one for line features. That data is parsed into one or more feature layers, according to the coverage/shapefile model. Polygons are attributed before the data enters the QC process. Process steps labeled as “Manual Process” require manual creation, editing or review of data. Note that in this schematic, only coverages are used in the digitizing process. Shapefiles can be substituted for coverages for manual capture and attribution tasks, however, all AML tools are written to run on coverages; shapefiles must be imported into coverage format for this reason.

The GRE’s current procedure for capturing hard-copy geologic map data involves digitizing all pertinent geologic features into two coverages, as defined by the coverage/shapefile data model. All line features are captured in a single arc/line coverage, whereas all points are captured in a single point coverage. This approach allows for topological relationships between features to be maintained in all data layers where they are coincident on the source map. Lines and points are then parsed to their respective data model layers (e.g., geologic contacts, faults, attitudes, etc.) using two AML scripts developed by the GRE—GENESIS.AML (lines) and CREATION.AML (points). Quality of captured data is ensured both manually and by employing AML scripts to check for data completeness according to the source map, and conformity to the data model.

Conversion of digital data is often more specialized depending on the source format, attribute structure, and the overall quality of the data. The GRE reviews digital data to assess the level of effort required to convert the data into the coverage/shapefile data model and to determine if additional editing will be necessary to bring the data up to GRE quality standards. Both quality and format are equally important and strictly enforced for all converted GIS data. Individual datasets (such as a single map) are converted to the coverage/shapefile data model manually. Multiple datasets (such as a collection of maps) from the same data model are often converted using AML scripts.

In order to ensure spatial coincidence between certain features such as faults and contacts, errors in existing data must be found and fixed. This can be problematic and time-consuming, considering that there are no readily available methods to ensure spatial coincidence between different coverages or shapefiles.

To supplement the GIS data, ancillary tables describing source map or source digital data and additional geologic unit information are generated. These tables are related to a data layer using a field common to both the ancillary table and the GIS data layer. Both shapefile and coverage formats do not allow for a permanent relationship between these tables and the GIS data; ancillary tables and GIS data must be temporarily joined when needed. Completed GIS data is combined with FGDC metadata, a Windows Helpfile containing source map information (legends, unit descriptions, cross-sections, etc.), and ArcView legend files to be used for symbolization, as part of the deliverable to each National Park.

## MIGRATION APPROACH

When revising the coverage/shapefile data model, an iterative approach was adopted. Each data layer in the coverage/shapefile model was reviewed to determine how each layer would be defined in a geodatabase. The features and functionality in a geodatabase were discussed with regards to how they could best be employed for a

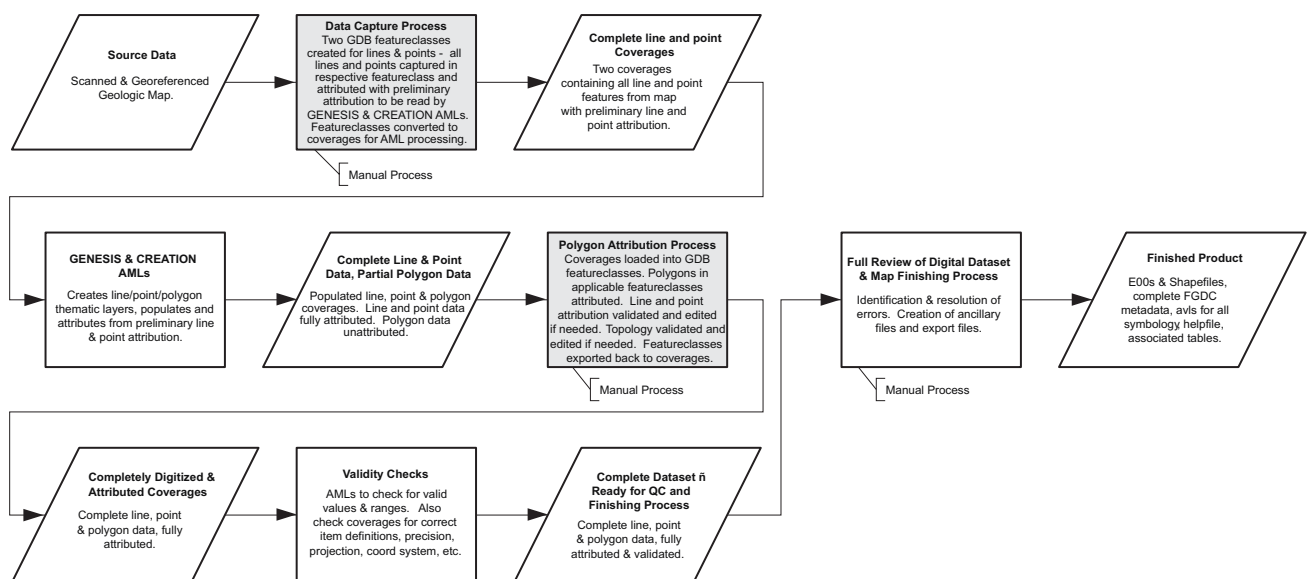
specific layer and its relationships with other layers. The resulting geodatabase schema, a detailed description of a data layer or layers, was implemented manually using dialogs and tools included in ArcCatalog. Data was loaded manually into the new schema and results were compared to data in the coverages/shapefile model. If any revisions were needed, the entire process was repeated until a final schema was agreed upon.

Implementing the data model in a geodatabase includes defining the data structure, setting up attribute value domains, and creating subtypes for use in topological validation. After evaluating Computer Aided Software Engineering (CASE) tools and other methods for implementing a data model, the GRE team decided to implement the geodatabase data model using an ESRI Developer Sample called Geodatabase Designer. The data model schema was stored in XML, as required by the Geodatabase Designer. The Geodatabase Designer is executed from ArcCatalog and, along with XML, provides the modular implementation necessary to load the layers needed for a specific geologic map. This is accomplished by designating an XML schema for each data layer in the geodatabase model. This aspect of modular implementation could not be accomplished using CASE tools, because they require one 'fixed' schema for all layers rather than individual schemas for specific layers. Although ESRI does not ensure long-term support for such a tool, the core functionality of the tool will most likely always work within the ArcCatalog architecture. Furthermore, ESRI supplies all source code for the Geo-

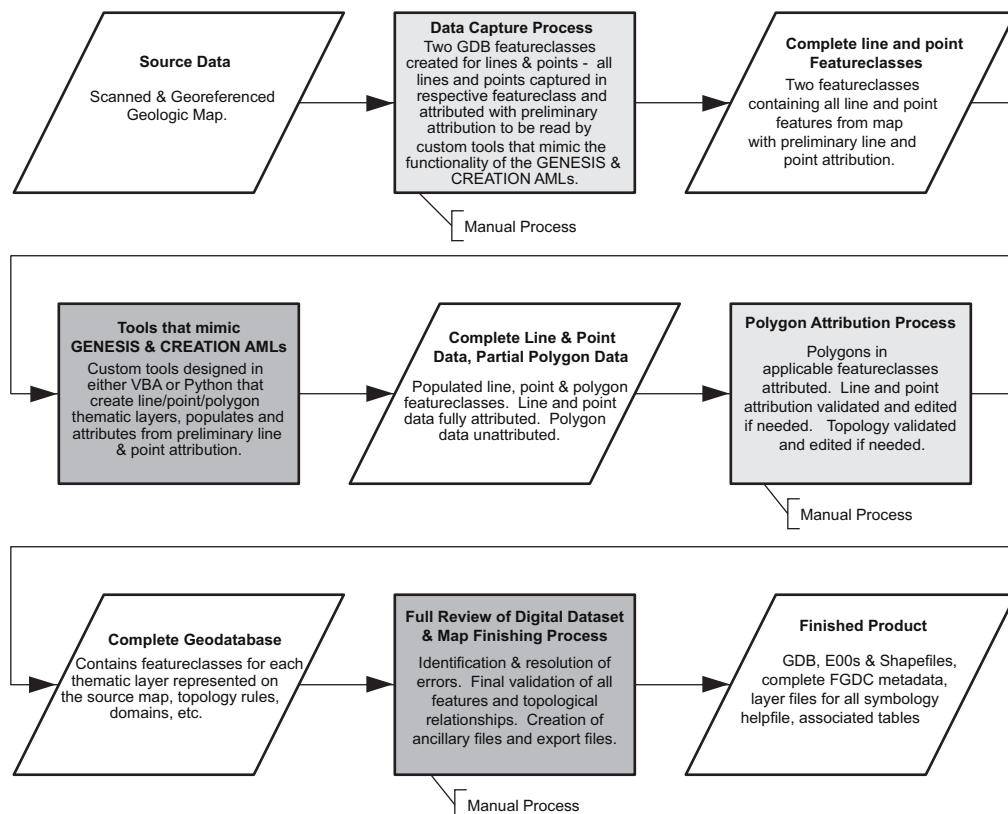
database Designer, thereby enabling users to modify and customize as needed.

It is of major importance to the GRE that production of digital geologic-GIS data not be interrupted by migration to the geodatabase data model. Equally important is the need to draw a well-defined distinction between data being produced in the old coverage/shapefile data model and data being produced in the new geodatabase data model. In order to reduce the impact of migration on geologic map production, this development work was done off-line until the new model had been designed, reviewed, revised, and released. This plan not only afforded time to properly develop the new model, but also enabled the continued use of AML scripts. Certain aspects of working in a geodatabase, such as topology validation, were inserted into the existing process to immediately improve the data being produced (Figure 2).

As part of the migration process, AML scripts that had previously been developed by the GRE to process data in coverage format must be redeveloped to work with data in a geodatabase. Currently, ESRI supports Python and COM languages, such as VBA, for automating processes within the ArcGIS framework. Any AML functionality that is not replaced by existing geodatabase functionality will have to be redeveloped using Python and/or COM. Until the new geodatabase model is implemented, however, our AML processes can remain in the digitization process alongside new geodatabase processes (Figure 2). Work has begun on using both Python and COM to replace AML processes (Figure 3).



**Figure 2.** Schematic showing NPS GRE modular process for creating GIS data in the coverage/shapefile data model with geodatabase processes in gray. These processes were implemented immediately to enable use of the geodatabase's domain, topology and attribute validation functionality. Export is still in coverage/shapefile data model.



**Figure 3.** Schematic showing NPS GRE modular process for creating GIS data in the geodatabase data model with projected script development in darker gray. Work has begun on replicating the functionality of the GENESIS and CREATION AML scripts within the new geodatabase framework. Note all process steps in this schematic produce or require geodatabase featureclasses instead of coverages. Also note removal of ‘Check Routines’ step—compensated for by inherent geodatabase functionality and XML schema-loading process. Export includes geodatabase, coverages, and shapefiles.

## CONCLUSION

The GRE recognizes the benefits offered by storing digital geology-GIS data in a personal geodatabase. The iterative process of revising data definitions to work within this new framework has begun, with good results. Topological validation has been successfully employed within existing digitization/conversion processes and new Python and COM scripts have begun to replace essential AML scripts. Conversion of legacy data to the new geodatabase data model will proceed, but only upon completion of digital geologic-GIS data for National Parks that have not yet been served by the GRE. The projected release date for the NPS Geology-GIS Geodatabase Data Model is the summer of 2005. The procedural approach developed for this process has worked well, with little to no interruption in production of GIS data.

## REFERENCES

O’Meara, Stephanie, Gregson, Joe, Poole, Anne, Mack, Greg, Stanton, Heather, and Chappell, Jim, 2005, National Park Service Geologic Resources Evaluation Geology-GIS Coverage/Shapefile Data Model, available at <http://science.nature.nps.gov/im/inventory/geology/GeologyGISDataModel.htm>.

## SOFTWARE REFERENCES

ArcGIS 8.3 and 9.0 (ArcCatalog, ArcMap, ArcEdit and Tables) – Environmental Systems Research Institute (ESRI) Inc., 380 New York St., Redlands, CA 92373, <http://www.esri.com>.  
Geodatabase Designer v2 – Developed by Richie Carmichael, Environmental Systems Research Institute (ESRI) Inc., 380 New York St., Redlands, CA 92373, <http://arcscrips.esri.com>.  
Microsoft Office Visio Professional 2003, Microsoft Corporation, <http://www.microsoft.com/>.